



LSST:UK DAC Roadmap and Integration Plan for DEV Activities

Work Package 2.1, Deliverable D2.1.1

Project Acronym LUSC-B
Project Title UK Involvement in the Legacy Survey of Space and Time
Document Number LUSC-B-03

| | |
|---|--|
| Submission date | 26/MAY/20 |
| Version | 1.1 |
| Status | Final |
| Author(s) inc. institutional affiliation | George Beckett (Edinburgh) |
| Reviewer(s) | Graham Smith (Birmingham); Tim Naylor (Exeter) |

| | |
|----------------------------|--|
| Dissemination level | |
| Public | |

Version History

| Version | Date | Comments, Changes, Status | Authors, contributors, reviewers |
|----------------|-------------|--|---|
| 0.1 | 24/FEB/20 | Defined high-level structure | MGB |
| 0.2 | 11/MAR/20 | Development of DAC effort expenditure and DAC Roadmap | MGB |
| 0.3 | 16/MAR/20 | Revisions based on feedback from RGM | MGB |
| 0.4 | 26/MAR/20 | Further Revisions based on feedback at all-team meeting | MGB |
| 0.5 | 30/MAR/20 | Document restructured | MGB |
| 0.6 | 2/APR/20 | Enhancements to the high-level DAC architecture | MGB |
| 1.0 | 27/MAY/20 | Updated in light of feedback from reviewers | MGB |
| 1.1 | 29/JAN/21 | Dissemination level changed from Project-internal to Public as agreed with Project Lead and Management | TMS |

Table of Contents

| | |
|--|-----------|
| VERSION HISTORY | 2 |
| 1 EXECUTIVE SUMMARY | 4 |
| 2 INTRODUCTION | 5 |
| 2.1 HIGH-LEVEL VIEW OF THE UK DATA ACCESS CENTRE | 5 |
| 2.2 GLOSSARY OF ACRONYMS..... | 7 |
| 3 UK DAC ROADMAP | 9 |
| 4 INFRASTRUCTURE PROVISION | 11 |
| 5 DAC-DEV INTEGRATION PLANNING | 13 |
| 5.1 PROJECT SUPPORT FOR DAC-DEV COORDINATION | 13 |
| 5.2 INTEGRATION OF DEV PRODUCTS INTO UK DAC | 14 |
| 6 TRANSITION TO PHASE C | 16 |
| 7 REFERENCES | 17 |

Index of Tables

| | |
|--|----|
| Table 1: UK DAC Roadmap for Phase B..... | 10 |
|--|----|

1 Executive Summary

Phase B is a critical time for the LSST:UK Consortium, as it an ideal opportunity for the UK astronomy community to prepare for the beginning of Rubin Observatory operations and to be ready to fully exploit UK data rights to deliver world-class scientific research.

The LSST:UK Consortium will deploy a UK-based Data Access Centre to not only serve Rubin Observatory data products to UK-based astronomers, but also to deliver enhanced capabilities (bespoke software and datasets), provided by Phase B DEV teams, to support the UK priority science areas.

The role of the DAC team is critical to the Consortium's preparations and, to this end, this document address three key threads of activity:

- The DAC team will deploy and maintain a prototype DAC infrastructure throughout Phase B, providing a preview of the capabilities that will be available to UK-based astronomers during operations, provide a training platform on which research groups can develop the skills and experience they need to fully exploit LSST data, and to enable development, validation, and integration activities for DEV teams.
- The DAC team will work with the main infrastructure provider, IRIS, to ensure the DAC and DEV activities have access to the infrastructure they need, and that IRIS provisioning plans are tailored to the requirements of the UK DAC during commissioning and early operations.
- The DAC team will work directly with DEV teams to ensure they have access to the data and services they need and to ensure DEV activities create products that can be integrated into the UK DAC and made available to end-users during Early Operations and, where appropriate, Commissioning.

2 Introduction

The LSST:UK Consortium is believed to be unique in the Vera Rubin Observatory (Rubin) ecosystem, in that it has funding for a number of research and development activities (called DEV activities, in the LSST:UK Long-term Plan [3]) in the run up to telescope operations. The funded DEV activities will typically enhance or add to the baseline Rubin Observatory data-products and data-analysis tools, targeting products and capabilities of particular relevance to UK astronomy and to International Science Collaborations' ambitions.

The Consortium has endeavoured to engage early on and actively with key stakeholders, such as:

- the LSST Project team and Rubin Observatory International Science Collaborations
- infrastructure providers (primarily, STFC IRIS)
- peer experiments such as Gaia and SKA

Over the course of Phase A, LSST:UK established a meaningful presence in key LSST bodies, including the Science Advisory Committee, Science Collaborations' leadership teams, and Rubin Observatory (at the time, called LSST) technical working groups. Through this presence, the Consortium has built up a strong relationship with the Rubin development staff, in the Data Management (DM) team and at stakeholder institutions such as IN2P3, in France, and NCSA, in the USA.

However, given that the Rubin Observatory is still more than two years from the beginning of operations and that the details of how it will operate are still being finalised, there is some risk to realising the benefits of LSST:UK's preparations, due to uncertainties in:

1. The software and services that the DM team will produce, including the pipeline software to create image products, the Data Access Centre (DAC) components that enable astronomers to use data, and even alignment to standards and community conventions, to enable integration with third-party tools and services.
2. The precise format of data products that will be produced from the survey (such as, details of cadence, depth, and format of products, as well as details of raw, source data from which products are derived).
3. How infrastructure will be provisioned in the UK, to host the LSST:UK Data Access Centre (DAC) through which UK-based data rights holders are likely to exploit LSST.

The science-related aspect of these uncertainties is best addressed through engagement between each particular DEV activity and the relevant international Science Collaborations. They are not discussed further in this report.

Technology-related elements of these risks (including data formats, software and interfaces, and infrastructure) are best addressed by a combination of the LSST:UK Data Access Centre (DAC) and DEV teams, and a strategy for doing this is a key element of this document.

2.1 High-level View of the UK Data Access Centre

The data to be produced by the Rubin Observatory represents a step-change in astronomy, with a volume and complexity that is not tractable using traditional workstation-based scientific analysis. Because of this, the Rubin Observatory is developing a suite of software services called the LSST Science Platform (LSP) to be deployed at two Rubin Observatory DACs, in USA and Chile, respectively. The two Rubin

Observatory DACs will be the primary route to access survey data for most US- and Chile-based astronomers. Further, those in the international community, who hold data rights, have an option to access data via the Observatory DACs, (at the time of writing) in return for an annual access charge.

The scale and ambition of the LSST:UK vision for exploiting LSST data has led the Consortium down a different path to many other international collaborations, in that LSST:UK will set up another DAC, modelled closely on the Rubin Observatory design, though deployed on UKRI infrastructure and tailored to specific UK astronomer science drivers, which are described in [3]. UK-based astronomers, who have data rights, are anticipated to typically access Rubin Observatory data products via the UK DAC, rather than the Project DACs. This means astronomers will be able to exploit the additional functionalities and data products being developed by DEV activities, and to use third-party surveys to which UK astronomers have access (and which US- and Chile-based astronomers may not have access).

In Figure 1, we provide a high-level overview of the expected composition of the UK DAC. To aid the plan described in this report, we highlight a partition of the components of the DAC into two parts: one with astronomer-facing elements, and one with elements to support DEV products to deliver added-value functionality and datasets (to astronomers through the UK DAC).

The astronomers' view will be based largely on software and services developed by the LSST Data Management Team (coloured in light blue):

- At the time of writing, three different user interfaces are anticipated:
 - A web-based view, called Firefly.
 - A scripting (Jupyter Notebooks) environment, called Nublado.
 - A programmatic API to undertake large-scale data analysis campaigns from remote, high-performance and high-throughput computing resources.
- Scientific data to be served to users is organised into three categories:
 - LSST Data Release Products, as defined in [2];
 - Third-party datasets from Ancillary Surveys, which are scientifically useful in combination with LSST data;
 - User-generated Products; for example, created using DEV-developed products.
- LSST Data Release Products, User-generated Products and possibly Ancillary Surveys are likely to be restricted to Rubin Observatory data-rights holders, details of whom will be served from the Rubin Observatory, though enforced locally within the UK DAC.
- The UK aspires to be an LSST Community Broker (coloured orange) and to provide access to the Observatory's nightly alert stream via the Lasair web and notebook services. At this time, it is not decided whether Lasair interfaces will be separate from or integrated with the three user interfaces (Firefly, Nublado, and API) noted above.
- Data will be ingested into the UK DAC from:
 - the Rubin Observatory, in the case of LSST Data Release Products and some externally generated User-generated Products.
 - Other surveys, in the case of Ancillary Surveys (coloured in dark blue).
 - Back-end DAC infrastructure in the case of UK-led User-generated Products—typically using software (algorithms and workflows) developed in DEV work packages (coloured green).

The make-up of the astronomer-facing elements is reasonably well understood and unlikely to change significantly between the time of writing and early operations. Conversely, the make-up of the DEV interface to the DAC (that is, elements coloured in

green) will vary from work package to work package: the illustration is a sample of one possible form, intended to highlight the following considerations:

- DEV algorithms and workflows that require access to large parts of an LSST Data Release and/ or third-party catalogues may benefit from having access to dedicated copies of that data, in a form most amenable to ingestion into the DEV workflow (which is not necessary the form used to serve the data to data-rights holders). Not only might this reduce the latency in delivering the associated User-generated Products, but it will also reduce demand on user-facing services and eliminate potential contention for access.

In contrast, where there is only modest demand on input sources – both LSST Data Release Products and third-party surveys – it is reasonable to expect this demand to be served by existing user interfaces (for example, the batch processing APIs for LSST data) without the need to set up separate copies of that data.

- DEV algorithms and workflows are likely to run on reserved “back-office” UK DAC infrastructure.

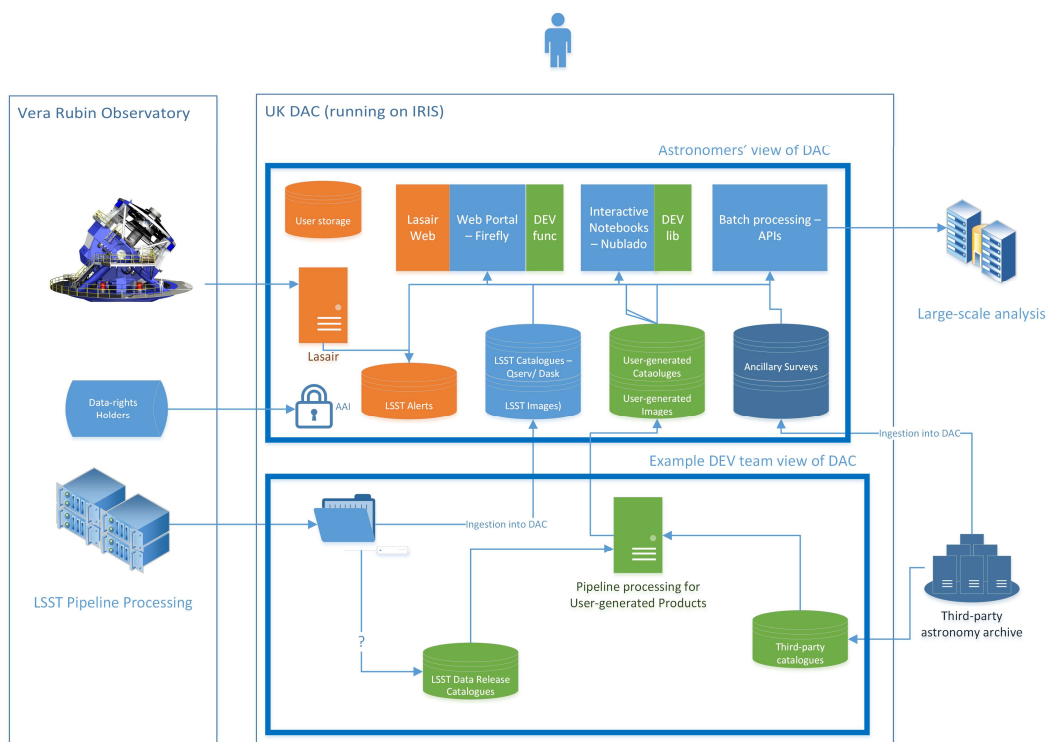


Figure 1: High-level View of UK Data Access Centre

2.2 Glossary of Acronyms

| | |
|------|---|
| DAC | Data Access Centre (or UK Data Access Centre work packages) |
| DEV | Development work packages |
| DM | LSST Data Management Team |
| LSP | LSST Science Platform |
| LSST | Legacy Survey of Space and Time |

LSST:UK DAC ROADMAP AND INTEGRATION PLAN FOR DEV ACTIVITIES

RSAP (IRIS) Resource Scrutiny and Allocation Panel

SC LSST International Science Collaboration

3 UK DAC Roadmap

A roadmap has been developed to deploy, maintain and support a prototype UK DAC platform throughout Phase B, informed by:

- The DM team’s plan for delivery of LSP services and functionalities
- The Rubin Observatory’s timeline for providing (commissioning-related) data previews.
- The LSST:UK Phase B (DAC and DEV) priorities and plans.

This will help UK-based astronomers to prepare for Rubin operations and, in the shorter term, will help with the integration of DEV products into the UK DAC.

The roadmap, which is to be finalised during 2020, is available on the LSST:UK Confluence site [4], though key points are summarised in Table 1, noting the following considerations:

- Target delivery date—the aim is to provide regular updates to the UK DAC platform, based on the forecast availability of LSST data previews and software, and aligned to the requirements of DEV activities.
- Platform—we expect the UK DAC to be hosted on infrastructure provisioned by IRIS.
- Functionality—the high-level service elements we expect to be available in each release.
- Data—the ancillary, user-generated, and LSST datasets that we intend to provide to aid end-user evaluation of each release.
- Access—the intended access mechanism.
- Helpdesk—the primary user-support mechanism that will be available.

One should note that the ancillary datasets listed in the roadmap are those that are proposed to be exposed to end-user astronomers, for evaluation of and familiarisation with the LSP and its capabilities: ancillary datasets that are required to support DEV activities will be documented in interface requirement documents (see Section 5) and included in the roadmap subsequently.

During Phase B, the DAC team will work with DEV activities to identify and deploy LSST (or third-party) technologies and data on which each DEV activity depends, plus to identify integration points at which DEV products can be incorporated into the platform.

LSST:UK DAC ROADMAP AND INTEGRATION PLAN FOR DEV ACTIVITIES

| | UKDAC0 | UKDAC1 | UKDAC2 | UKDAC3 |
|----------------------|--|--|--|---|
| Delivery Date | 2020-05-01 | 2020-10-01 | 2021-12-01 | 2022-12-01 |
| Platform | IRIS | IRIS | IRIS | IRIS |
| Functionality | <ul style="list-style-type: none"> LSP components: Firefly, Nublado Non-LSP components: none | <ul style="list-style-type: none"> LSP components: Firefly, Nublado, qserv(TBC) Non-LSP components: Lasair-ZTF (TBC), basic user storage | <ul style="list-style-type: none"> LSP components: Firefly, Nublado, qserv Non-LSP components: Lasair-ZTF, user storage, batch compute to support DEV and Commissioning activities | <ul style="list-style-type: none"> LSP components: all Non-LSP components: Lasair, user storage, batch compute, DEV services (TBD) |
| Data | UKIDSS DR11 (TBC) via existing WFAU TAP (catalogues) and SIA (images) services | UKDAC0 holdings, plus PS1 (catalogues) via DAC-hosted TAP and SIA services. Lasair-ZTF and other prioritised ancillary datasets (TBC). | UKDAC1 holdings, plus all ancillary data (images & catalogues) needed for Commissioning. Data Preview 0 datasets (DESC DC2, HSC) | UKDAC2 holdings, plus Data Preview 2 (LSSTCam Science Validation Surveys), all ancillary data needed for start of main survey operations |
| Access | <ul style="list-style-type: none"> Authentication: EGI check-in Accounts: admin, internal DAC test users | <ul style="list-style-type: none"> Authentication: EGI check-in Accounts: admin, invited test users from LSST:UK Consortium | <ul style="list-style-type: none"> Authentication: Project DAC authentication (TBC) Accounts: admin, invited users involved in Commissioning, DEV WPs and other activities (TBD). | <ul style="list-style-type: none"> Authentication: Project DAC authentication Accounts: admin, all UK data rights holders (respecting any restrictions on access to Commissioning data) |
| Documentation | Only as provided by Project | Project-provided, plus online documentation for invited test users | Project-provided, plus UK-specific online documentation to support Commissioning activities | Project-provided, supplemented by UK-specific online documentation to support authorised users |
| Helpdesk | GitHub Issues (internal DAC only) | GitHub Issues (test users) | Project-provided or UKDAC helpdesk | Project-provided or UKDAC helpdesk |

Table 1: UK DAC Roadmap for Phase B.

4 Infrastructure Provision

During Phase A, it became clear that the infrastructure needed to support LSST:UK science would most likely be provided by the shared-infrastructure project IRIS [www.iris.ac.uk].

At the time of writing, IRIS offers three kinds of infrastructure:

- Large-scale high-throughput computing and storage on grid infrastructure managed by GridPP [www.gridpp.ac.uk].
- High-performance computing resources from DiRAC [www.dirac.ac.uk].
- Cloud computing and storage [available from several IRIS member institutions].

Time and capacity on these different infrastructures is secured through an annually (April—March) Resource Scrutiny and Allocation Panel (RSAP). The RSAP treats LSST:UK as a single user and expects the consortium to formulate a single unified request to IRIS, late in each calendar year, based on the aggregated requirements of the LSST:UK members.

For LSST:UK, the DAC team is the point of contact with the RSAP. The DAC team has engaged successfully with the RSAP since its inception in late 2018 and has, at the time of writing, secured more than 25 million core hours of computing time and 2 Petabytes of working storage.

As well as applying for infrastructure on behalf of the Consortium, the DAC team also works with IRIS to ensure that the infrastructure procured by IRIS is relevant to LSST:UK's needs. The DAC team maintains a five-year forecast of infrastructure requirements, which feeds into the IRIS Delivery Board's provisioning plans, and highlights potentially demanding LSST:UK use cases to the IRIS Technical Working Group, to ensure that LSST:UK technology needs are considered when planning and provisioning new infrastructure.

Around October of each year, the DAC team requests details, from both the DEV activities and the wider UK Consortium, of forthcoming infrastructure requirements. Given the potential complexity of the infrastructure available, the DAC team works with DEV teams and other Consortium activities to understand and properly size their requirements, so that a coherent and strong bid for resources can be submitted to IRIS for the December deadline. As an allocation process (rather than an award process), it is common for the RSAP to engage with the DAC team during application preparation.

Even after resources have been awarded, the DAC team continues to be a liaison between IRIS and the Consortium, monitoring and validating any deviations from the planned usage by LSST:UK, lobbying to IRIS for resolution of infrastructure issues, and securing from IRIS any top-ups or changes to the allocation that may be required.

As Phase B progresses, it is anticipated that the type and level of demand LSST:UK will place on IRIS will grow. The requirements that are envisaged to be the most demanding for IRIS are:

- Providing capabilities for LSST:UK to operating significantly distributed, multi-Petabyte databases.
- Providing low-latency infrastructure for consuming and processing the LSST alert stream in pseudo-real-time.
- Providing rapidly scalable cloud resources to adapt to the variable demand from users for analysis resources over the lifetime of each LSST Data Release.

LSST:UK DAC ROADMAP AND INTEGRATION PLAN FOR DEV ACTIVITIES

The DAC team will continue to work with IRIS to satisfy these demands, along with others identified through the work of the DEV work packages.

5 DAC-DEV Integration Planning

At the time of writing, six DEV work packages are active (from the Phase B programme):

- WP3.2: LASAIR—the UK transient broker for LSST
- WP3.5: LSST and near infra-red data fusion
- WP3.7: Low-surface-brightness science using LSST
- WP3.9: LSST Point Spread Function, sensor characterisation and modelling
- WP3.10: UK Contributions to DESC Operations
- WP3.11: Cross matching and astrometry at LSST depths

These work packages will produce software to:

- Enhance the DAC user-facing functionality.
- Enable the creation of User-generated Products, to supplement the baseline LSST Data Release Products.
- Support users to incorporate ancillary datasets, alongside LSST Data Products and User-generated Products [2], into their research.

To maximise the impact of outputs from the DEV work packages, careful consideration needs to be given to how DEV outputs are integrated into the UK DAC, including:

- How DEV activities' software is integrated with software and services provided by the LSST DM team and by the UK DAC team.
- How enhanced functionalities are made available to users, via the UK DAC.
- How User-generated Products and ancillary surveys are made available in the UK DAC.
- How the UK DAC infrastructure and services are configured to maximise the impact of products from the DEV work packages.

This consideration is best addressed by the DAC team and each DEV activity, in collaboration.

5.1 Project Support for DAC-DEV Coordination

As part of the setup of Phase B, an enhanced project-support mechanism has been devised, building on experience from Phase A and recognising the need for the outputs of different work packages to be integrated together before the end of LSST Commissioning Phase. These mechanisms are described in detail in the Project Management Plan [1], though summarised here.

- A communication plan has been developed to encourage regular and frequent communications within the UK team. In particular, the whole team will meet (face-to-face or virtually) at least every three months, to coordinate cross-project activities, to update plans and to disseminate progress. In addition, several mailing lists have been created for day-to-day information sharing, and a Slack workspace has been set up to encourage more spontaneous and interactive discussions.
- Collaboration – building on good practice from other distributed development projects, various tools have been adopted to streamline and support collaboration. Most notably, there is a well-established project routine built around Atlassian Confluence and, new to Phase B, GitHub. All project documents and materials are stored in Confluence [<https://lsst-uk.atlassian.net/wiki>] and software development is coordinated in GitHub [<https://github.com/lsst-uk>].
- Community engagement – mirroring the philosophy of LSST, the UK consortium has adopted an Open Source strategy for its work. Almost all of the Confluence and GitHub materials are visible to the whole collaboration, the project team regularly

updates the community on its plans and progress, and software and data is released (where permitted) as Open Source.

- Stakeholder engagement – to be successful, the Consortium needs to actively engage with key stakeholders in a number of fields: the LSST:UK leadership team is represented on the STFC IRIS Delivery Board and Technical Working Group; the Consortium has an on-going dialogue with peer projects, such as Gaia, Euclid and SKA; and the DAC team are working collaboratively with other Rubin stakeholders including LSST:France, the Dark Energy Science Collaboration, and the LSST Alert team.

5.2 Integration of DEV Products into UK DAC

All but one of the Phase B DEV work packages will deliver products (software or datasets) that need to be integrated into the UK DAC. Funding has been identified, in the Phase B award, to support DAC engagement with each individual DEV work package, and four of the six DEV work packages (specifically WP3.5, WP3.7, WP3.10, and WP3.11) are expected to call on this effort, since:

- WP3.2, which considers the scientific exploitation of the LSST alert stream, has a dedicated DAC complement (WP2.3), which will address the integration of alert-stream services into the UK DAC platform and infrastructure;
- WP3.9, which considers the sensor characterisation modelling of the point spread function for the LSST camera, will not directly contribute functionality nor datasets into the UK DAC.

DAC engagement with the DEV work packages is an on-going activity, though is particularly important at two stages:

- Early on in the timeline of a DEV work package, to agree on how the DEV work package's products integrate into the DAC, and to agree the interfaces that should be employed between the DAC platform and the DEV products.
- Towards the end of each DEV work package, to validate the DEV outputs, finalise the integration of these outputs into the DAC platform, and to agree an operating model for DEV outputs in LSST:UK Phase C.

Agreeing and defining the form of the DEV view, for each of the four identified DEV work packages, is the crucial task for the initial DAC interaction with each work package, along with an understanding of what technologies will support the interactions between the DEV view and the astronomer view.

The findings of this work will¹ be documented in an *Interface Requirements Document*—one for each DEV work package—which will include, at least:

- Details of the datasets that are required by the DEV team, along with when the DEV team need them and how the DEV team will use them.
- Details as to how the DEV outputs will integrate into the UK DAC, and the interfaces that need to be provided by the DAC team for this.
- The interfaces and standards the DEV team need to adopt and develop against to make integration easy.
- The approach to validate DEV team's products in the DAC, within the scope of the Phase B effort (effectively DEV commissioning).

¹ At the time of writing, work on Interface Requirements Documents has already begun.

- The infrastructure and support requirements, in production, to operate the DEV team's products.

6 Transition to Phase C

The operational plans described in this report are intended to realise a working UK DAC towards the end of Phase B, which integrates functionality and user-generated-product workflows from DEV activities, to enhance the capabilities for UK-based astronomers. The plan should also generate forecast requirements for infrastructure, which will then be shared with the IRIS Delivery Board to influence their provisioning plans into Phase C.

Phase C, as described in [3], focuses on scientific exploitation of the early LSST Data Releases. In that phase, the primary functions of the DAC team will change to be a service-operation role managing access to available (IRIS) infrastructure and supporting UK-based astronomers to use the Data Access Centre; and operating/ refining the DEV workflows to produce UK-led User-generated products for the community.

7 References

- [1] G. Beckett, LSST:UK Science Centre Phase B Project Management Plan, Version 1.6 (February 2020)
- [2] LSE-163 – LSST Data Products Definition Document
- [3] R.G. Mann, LSST:UK Long-Term Plan, Release 1.1 (May 2016)
- [4] LSST:UK DAC Roadmap, <https://lsst-uk.atlassian.net/wiki/spaces/LUSC/pages/921337857/Phase+B+DAC+Roadmap>, (accessed 25th May 2020)